Ethics of artificial intelligence and health care

Professor Tim Dare
BA, LLB(Hons), M.Jur (Dist), PhD
Professor of Philosophy
Faculty of Arts
University of Auckland

Artificial intelligence (AI) aims to mimic and improve on some human cognitive functions. Humans can identify patterns, apply rules, classify data, and make predictions and decisions based on those activities. Such activity is central to medical practice. Diagnostic radiologists, for instance, examine medical images to identify signs of pathology. The expert radiologist draws on training and experience to identify features in the images that match those seen in cases that have proven to be pathological. Likewise, general practitioners (GPs) assessing the likelihood that a patient has some condition draw on their training and experience to decide whether the patient has features known to be characteristic – or symptomatic – of that condition. Health care policy makers and administrators bring similar cognitive skills to bear when making decisions about population level health needs and more immediate resource and staffing allocations: what happened to June–July hospital admission rates the last time April flu rates looked as they do this year?; what health needs can we predict over the next five years given what information we have about the population our system is serving?

The datasets that would inform these processes in an ideal world are huge. There are too many cases; too many images; too much research; too many variables affecting admission rates; too many combinations between variables; and so on for humans to identify and process. Much of that data now exists in electronic form, or in forms that can be accessed electronically by natural language processing systems. Some clinical and health data may be collected 'manually' as researchers, health administrators, clinical staff, and others enter health information onto computers, GPs claim subsidies for patients, patient appointments are entered as Accident Compensation Corporation claims, or the like. Other data is electronic from the outset: images, input from health-care devices that create digital records as they weigh, ventilate, and pump. The intensive care unit ventilators, health apps on mobile phones, a GP's digital thermometer, blood pressure machines, and scales can all generate electronic records that could be aggregated into datasets. As the 'internet of things', and in particular the internet of medical things expands, so too does the list of potential sources of health data.

These electronic datasets create the opportunity for computers to at least enhance, and perhaps take over, many of the reasoning tasks previously carried out by humans. Computers can access and process vastly larger datasets than their human counterparts. They can identify patterns indiscernible to humans without tiring, and without running out of capacity to consider more cases. It is tempting (and true) to say that they can do so more quickly than humans, but reference to their speed misses the point: humans simply could not get through the data processing tasks managed by computers. So while it is true that computers are fast, their speed is part of their capacity to process vast searchable datasets at the outset, rather than a separate feature.

In the early days, computers simply ran algorithms – a problem solving process or set of rules – set by programmers. Algorithms can be very simple, perhaps a straightforward 'if x then y' rule, or very complicated, involving multiple steps and complex mathematical formulas. Simple versions may look very much like equally simple algorithms used by humans. For example, if my GP is considering recommending a prostate-specific antigen test for me, they are likely to work their way through a checklist – a nonautomated algorithm of sorts: is my patient male? If yes, is he over 50? If no, is he over 40 with a family history of prostate cancer? Is he urinating frequently? And so on. It is easy to imagine a computer running through a similar checklist and making recommendations, though perhaps it is not obvious what advantage there would be to delegating such task.

We might have reason to do so if we feared that the risk factors for prostate cancer were much more complex than our simple algorithm assumed. The number of potential predictor variables in electronic health records may be enormous and the combinatorial possibilities unimaginably large. We might proceed by choosing a limited number of commonly collected variables, but we would risk locking ourselves into the short-sightedness we are attempting to address; the problem might be with our choice of variables and not just with the reliability of processing them.

Suppose then we give computers access to all the electronic data we have about patients who have been accurately diagnosed with prostate cancer and set the computer the task of identifying correlations between the data and the diagnosis? The computer could look at vast numbers of cases and vast numbers of predictor variables and combinations between them, and identify correlations that humans have missed, perhaps because the correlations were only apparent across very large datasets, sets too big for humans to manage, or perhaps because the correlations hold between disease status and complex combinations of variables. And we might go further. The computer could 'learn' from its own outputs. Suppose, given ongoing access to diagnostic outcomes, it notices that risk assessments it had generated on the basis of some correlations were less reliable than it had initially indicated – perhaps its early predictions contained more false positives than would have been the case had it relied on different correlations or assigned different weight to variables. It then adjusts its own algorithms accordingly. Now the computer would be learning – machine learning – from the data, creating its own algorithms, rather than simply relying on those set for it by its human designers. We might regard it as exercising AI.

It has been shown that AI, more or less as described here, can operate in health care and can at least match humans. A 2018 paper reports a study in which researchers fed de-identified data on hundreds of thousands of patients into a series of machine learning algorithms powered by Google's massive computing resources.[1] The algorithms were able to predict and diagnose diseases, from cardiovascular illnesses to cancer, and predict related things such as the likelihood of death, the length of hospital stay, and the chance of hospital readmission. Within 24 hours of a patient's hospitalisation, for example, the algorithms were able to predict with over 90% accuracy the patient's risk of dying. Earlier, the same team used data on eye scans from over 125,000 patients to build an algorithm that could detect retinopathy, the number one cause of blindness in some parts of the world, with over 90% accuracy, which is on par with board-certified ophthalmologists.[2] Going back to our simple prostate cancer example, a number of studies have shown the potential for AI to improve diagnosis and the identification of treatment options for the disease.[3,4] Of course, not all of the news about AI, in health care and beyond, has been so positive. It is widely accepted, even by those who support the introduction of AI, that the technology promises significant ethical and legal challenges. According to recent books, algorithms are 'weapons of math destruction' increasing inequality and threatening democracy;[12] automated decision-making tools 'profile, police, and punish the poor';[13] tech products are 'full of blind spots, biases, and outright ethical blunders' which 'exacerbate unfairness and leave vulnerable people out'.[14]

Some of these challenges may seem especially pressing in health contexts. Consider the fundamental concern in medical ethics to treat patients with respect, a concern that underpins the obligation to provide patients with full information and to obtain consent in almost all cases (See the Code of Health and Disability Services Consumers' Rights, especially rights one (right to be treated with respect), six (right to be fully informed), and seven (right to make an informed choice and give informed consent). The use of AI may make it difficult to meet these obligations, at least as they have been traditionally understood. It may not be possible, for instance, for humans to explain, or even to know, why a complex machine learning system has classified a case one way rather than another. The classification may rest on complex correlations that cannot be reverse engineered. Algorithms, that is, may not be transparent or scrutable: they might be black boxes.

Some regulation of the use of AI has gone a way toward banning such systems. Under new European data protection guidelines, those affected by automated decision making systems are entitled to 'meaningful information about the logic involved'.[5] Our own Privacy Commissioner and Chief Government Data Steward have issued a set of principles for the use of data and analytics, which specify that 'explanations of decisions – and the analytical activities behind them – should be in clear, simple, easy-to-understand language'.[6]

But, I have argued that the demand for explainable AI (in health and elsewhere) is mistaken.[7] Health professionals do not, and cannot, explain how a lot of familiar health technology works – digital thermometers; magnetic resonance imaging scanners (MRIs)? These familiar tools are neither transparent nor explainable (MRIs rely on quantum mechanical explanations of the spin and orbital angular momentum of subatomic particles, and 'I think I can safely say that nobody understands quantum mechanics').[8] But patients should not care. What matters is not transparency, or 'explainability', but whether there is evidence of reliability: it does not matter how the thermometer identifies my temperature as 36.7°C, providing that I know that it does so reliably. It is evidence of reliability – rather than transparency – that we should insist on in the case of automated decision-making systems too. Evidence of reliability – rather than an explanation of how technology works – also seem to meet the Code of Health and Disability Services Consumers' Rights, right six, to be 'the information that a reasonable consumer, in that consumer's circumstances, would expect to receive'. When I ask about the MRI my GP will probably give me evidence that the scans are accurate and useful, and that – rather than a course in quantum mechanics – seems just the sort of thing I am likely to want.

AI also raises important questions about our privacy and consent, at least as those interests are currently understood.

Consent is widely regarded as essential for legitimate access to and use of health information. Again, it is an important aspect of respecting persons, but our understanding of consent and its importance was forged when information was gathered and aggregated in clear transactions, and in ways that allowed us to track its use toward clearly-articulated goals. In an era of vast datasets in which end-uses and users are often unclear at the collection point, and in which data will be combined, reprocessed, and reused in ways that make it difficult to establish straightforward relationships between providers, processors, and users, it is unclear how traditionally-understood consent might work. Even where it is possible to seek informed consent, the size of datasets may make it prohibitively expensive. Some of our concerns might be met by limiting the use of 'unconsented' data to de-identified datasets, but many important applications require identification. This is not to say that AI requires us to abandon consent. We do need to be clear, however, what holding on to the traditional consent paradigm will cost in terms of the forgone advantages of at least some uses of AI.

Privacy has become a flagship right – we have Privacy Acts, Officers, and Commissioners. We certainly think we have moral rights to privacy (and that they are everywhere under threat). It is certainly true that AI threatens our interests in privacy as traditionally understood. In a famous case, an algorithm allowed an American pharmacy chain to work out that a young woman was pregnant and send her (or, the detail that started the trouble, her father) coupons for baby goods before she had said anything to anyone.[9] Regulation of AI might address some of these problems, but, like our interest in consent, I suspect it would be a good thing if there were movement on both sides. On the one hand, we could limit the use of data to find out 'private things'. On the other, we could all recognise that our current concern with privacy is not always a good thing. Privacy has clear benefits – no one wants to be under constant surveillance – but it is often used to protect people against unjustifiable discrimination. Think about sexual orientation. When discrimination was likely to follow knowledge that a person was gay, people who identified as gay had good reason to keep their sexual orientation private. As we have adopted more sensible views about sexual orientation, privacy has become less important and the resulting openness has been a very good thing. We are all better off in a world in which we do not need privacy about sexual orientation. And it seems that at least some of our concern for privacy is relatively recent. When we lived in smaller communities – villages or small towns, or in rural districts served by phone systems that allowed others to know when we got a call (and perhaps even to listen in) – our neighbours were likely to know a good deal about us. Our concern for privacy is in part a consequence of the urbanisation that has made it possible for us to keep large parts of our lives secret. We have come to think of that secrecy as normal and important, but it is not clear we are right. Privacy may be corrosive and isolating. Knowing less about our neighbours means we do not know who needs a hand. We are more likely to feel threatened and alienated by those we do not know. Perhaps, properly regulated with respect to privacy, AI will allow us to reclaim some of the benefits of an earlier time.[10]

Another common concern about AI that may seem especially relevant in a health context concerns the role or opportunity for human judgment or oversight. Again, the General Data Protection Regulation gives those affected by automated decision-making systems a right, 'not to be subject to a decision based solely on automated processing'[5] and the New Zealand principles for the safe and effective use of data and analytics specify that, '[a]nalytical processes are a tool to inform human decision-making and should never entirely replace human oversight'.[6] As others have pointed out, the right poses little practical constraint – few systems do not, or cannot, or would not wish to, include a human in the loop at some point. The prostate algorithm may generate a risk score for me, but my GP will call me in to discuss its significance. Perhaps health resource allocation processes could be fully automated. But, there is some suggestion that restrictions on delegations of power in New Zealand prohibit delegation, other than to a person). Nonetheless, it is important to see that including humans in the loop is unlikely to improve the accuracy of algorithms. Machines are, or soon will be, more accurate at, for instance identifying and interpreting complex risk factors, than any of the alternatives available to us – most obviously relying on guided or unguided clinical judgment – and, furthermore, it is likely to be easier to state and measure (and remeasure) their accuracy more precisely than that of alternatives; we know how right or wrong they are and so can (try to) accommodate their error rates.

There is another aspect to the importance of human judgment, however, which might be especially significant in health contexts. Amazon has a 'chaotic storage algorithm', which tags every item entering its warehouse with a barcode and assigns it to a location based on available shelf space (i.e, not by type, or manufacturer, or alphabet, etc). There are no humans in the loop, but it doesn't seem to matter. We might not be so sanguine when AI is used in contexts in which relationships matter. Care providers relying on AI suggest Brent Mittelstadt and Luciano Floridi 'may be less able to demonstrate understanding, compassion and other desirable traits found within "good" medical interactions in addition to applying their knowledge of medicine to the patient's case. Put another way, the patient's body and voice may increasingly be replaced or supplemented by data representations of state of being if [AI] practices are adopted in medicine'.[11] But the conclusion seems too quick. Reliance on AI could reduce patients/clients to mere data, but surely it need not; AI might free health-care professionals to focus on relationships, handing time-consuming diagnostic tasks to systems that are better at some aspects of their current role than they are, and it might spawn new roles or aspects of roles focused on the caring aspects of the professions. It is important to remember that practices are not fixed; their identification with apparently defining goods may be contingent. As health providers and consumers come to appreciate the potential of AI to serve the central health-promoting functions of caring roles, they may come to understand the goods those roles deliver differently. That may be a lesson to be taken on board by those currently training for roles in the health-care system, and for those who are training them.

References

1. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med. 2018;1(1):18.

2. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA [Internet]. 2016 Dec 13;316(22):2402–10. Available from: https://dx.doi.org/10.1001/jama.2016.17216

3. Berglund E, Maaskola J, Schultz N, Friedrich S, Marklund M, Bergenstråhle J, et al. Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. Nat Commun. 2018;9(1):2419.

4. Nagpal K, Foote D, Liu Y, Wulczyn E, Tan F, Olson N, et al. Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. arXiv Prepr arXiv181106497. 2018;

5. General Data Protection Regulation 2018 (European Union).

6. Stats NZ. Principles for safe and effective use of data and analytics. 2018.

7. Dare T. Tread carefully with big data ethics. Newsroom [Internet]. 2018 Jul 11. Available from: https://www.newsroom.co.nz/2018/07/10/147233/tread-carefully-with-big-data-ethics?amp=1

8. Feynman R. The character of physical law. London: Cox and Wyman Ltd;1967.

9. Fry H. Hello world: how to be human in the age of the machine. Random House;2018.

10. Dare T. Tim Dare: privacy is not always a good thing. New Zealand Herald. 2017.

11. Mittelstadt BD, Floridi L. The ethics of big data: current and foreseeable issues in biomedical contexts. Sci Eng Ethics. 2016;22(2):303–41.

12. O'Neil C. Weapons of math destruction: how big data increases inequality and threatens democracy. Broadway Books;2017.

13. Eubanks V. Automating inequality. St. Martin's Press;2018.

14. Wachter-Boettcher S. Technically wrong: sexist apps, biased algorithms, and other threats of toxic tech. WW Norton & Company; 2017.